

Convolutional Neural Network を用いた電子顕微鏡連続切片画像 Z 補間

内橋 堅志^{*1} 石井 信^{*1} 大羽 成征^{*1}^{*1} 京都大学 大学院情報学研究科 システム科学専攻

Connectomics is a recent active field of research dedicated to the understanding of the map of neural connections in the brain (connectomes). One of the most formidable challenges in the field is the reconstruction of the 3D image of the connectomes from the stack of 2D EM slices. The most up-to-date solution to this problem requires human proofreading, and fully automated algorithms are yet to be developed. The aim of this research is to make a progress toward the full automation of the reconstruction. In the endeavor, we used deep convolutional network, an architecture known for its effectiveness in the interpolation of non linear structures, to interpolate the images between the 2D slices. Our method was able to show superiority over baseline methods in some situations.

1. 序論

コネクトミクスと呼ばれる新しい研究分野では、脳における神経細胞間の結合や神経繊維配線の構造を網羅的に調べ、その機能的意味を理解してゆくことを目的としている。この研究分野が生まれた背景には、重要なブレイクスルーが二つあった。一つは、これまでの技術では見ることが難しかった神経細胞の詳細な繊維状構造を電子顕微鏡によりナノメートルのスケールで見られるようになったことである。もう一つは、Automatic Tape-Collecting Lathe Ultramicrotome (ATLUM) という技術により、脳標本を数十ナノメートル間隔という薄さでスライシングした連続切片画像として自動的に得られるようになったことである [Hayworth 06]。このようにして得られたデータを電子顕微鏡連続切片画像と呼び、典型的には 1 ピクセルあたり 3 ナノメートル角程度の分解能による 1000 × 1000 ピクセル程度の画像 100 枚程度を 30 ナノメートル間隔で取得したものが 1 ブロック分の解析対象となる。

電子顕微鏡連続切片画像解析の第一目的は、ブロック内に含まれる神経細胞の詳細な三次元構造の再構成である。すなわち、各ピクセルを神経細胞を表すピクセルとそうでないものに分け、さらに神経細胞個体ごとに別のセグメントになるように分けることがゴールである。このために、Harvard 大学や MIT の研究グループでは段階を追った画像処理方法を採用している [Kaynig 13]。彼らは、各切片画像上のピクセルをパターン認識によって細胞膜上とそうでないものに分類し、膜で囲まれたブロップと呼ばれる三次元的な塊を巧みな画像処理によってまとめ、さらにブロップ同士を繋ぐことによって神経細胞の構造を決定する。しかし、後半の過程では専門家の手による校正によらねば十分な精度が得られないと考えられており、この校正に多大な労働力が必要となっている。

我々は、自動再構成の精度が上がらない原因が電子顕微鏡連続切片画像データにおける Z 方向の解像度 (30 ナノメートルの切片間隔) が X,Y 方向の解像度 (切片内ピクセルサイズ 3 ナノメートル) と比べて低い点にあると考える。Z 方向の構造連続性に関する情報が失われることによって、細胞膜の分類を行う際にも、ブロップを決定する際にも不利である。よって Z 方向の画像補間 (超解像) を行うことで、全過程の精度向上が期待できる。本研究において、我々は教師あり学習に基づく

Z 方向補間を行った。教師あり学習には、神経繊維画像の局所パターンを精度よく認識させるために Convolutional Neural Network(CNN)[LeCun 98] を用いた。提案手法の有効性を示すため、線形モデルを用いたより単純な方法と性能比較した。

2. 関連研究

電子顕微鏡連続切片画像データを取得して神経細胞およびその間の結合構造を三次元再構成する試みには近年注目が集まっている [Lichtman 08]。切片画像に含まれるニューロンを人力で一つずつ異なる色に色分けしてゆくためのソフトウェアが開発された [Helmstaedter 11]。さらに、色分けを補助するシステムの研究も行われた [Chklovskii 10]。連続切片画像におけるニューロンの色分け分類を CNN によって自動化する試み [Turaga 12] や、こうして色分けされた画像に対して三次元再構成を行う研究 [Funke 12] が行われている。しかしこれまでの方法による再構成精度は十分ではなく、さらに大規模なデータセットに対して処理時間がかかりすぎる問題があった。Harvard 大学のグループが処理時間の問題を解決しパイプライン化することに成功 [Kaynig 13] したものの、依然として人間による校正プロセスを必要としており、未だ完全な自動化には至っていない。

本研究で主に用いる CNN とは、各層で convolution と pooling を交互に行うことによって結合をスパースにした多層ニューラルネットワーク (DNN) の実装の一つである [LeCun 98]。画像認識の文脈から見た CNN の特徴は、画像データの特徴を表現するために人手で構成した特徴量抽出子を必要とせず、画像と対応する教師信号を与えれば DNN が持つ複数の中間層に従来の特徴量に対応する重みが自然に現れるという点にある。DNN の中でも層の多いモデルは、その応用に伴う困難として、誤差逆伝播法によって行われる勾配の学習がうまくいかない問題や、トレーニングエラーが下がっているにも関わらずテストエラーが下がらない過学習の問題、ハイパーパラメータチューニングの難しさ、計算時間の長さなどの問題があった。以上のような問題点は、LeCun らによる DNN を訓練する方法の提案 [LeCun 89] や、Hinton らによる新しい活性化関数の提案 [Nair 10]、さらに計算機の性能向上などにより部分的に改善され、CNN の有効性が高まった。

3. 手法

電子顕微鏡連続切片画像データの前処理と、Z 補間のための方法として比較した四つの手法について述べる。

連絡先: 内橋 堅志, 京都大学大学院情報学研究科システム科学専攻石井研究室, 〒 606-8501 京都市左京区吉田本町 36-1 工学部 1 号館, tel: 075-753-4908, E-mail: uchihashi-k@sys.i.kyoto-u.ac.jp

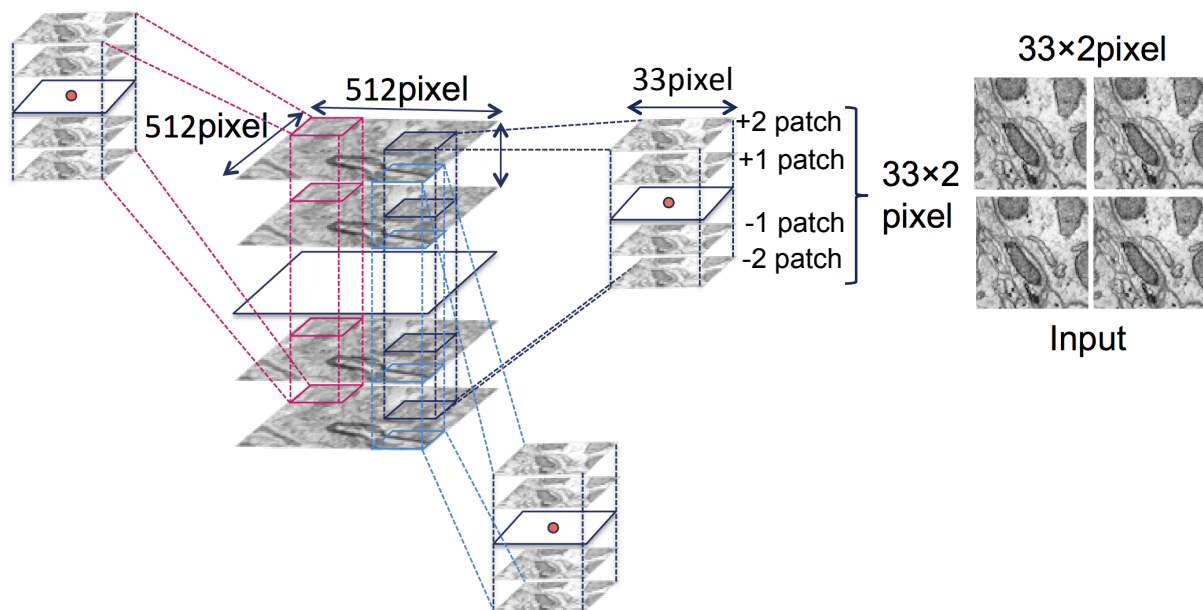


図 1: 推定対象切片 $Z = 0$ の上下連続 2 切片 $Z = +2, +1, -1, -2$ から、右端の入力パッチ画像を切り出すプロセスを示した。各切片から X 座標, Y 座標を共有するように 33×33 の領域を切り取って左上から $Z = +2, +1, -1, -2$ の切片を並べて入力画像パッチとした。この際、推定対象切片の中心ピクセル $(X, Y) = (17, 17)$ の輝度値を教師信号として取得した。

3.1 前処理

本研究では、Sebastian Seung らが公開している電子顕微鏡連続切片画像データ^{*1}を用いた。トレーニング用、テスト用の電子顕微鏡連続切片画像として与えられているそれぞれ 1024×1024 ピクセルの 100 切片ずつ合計 200 切片に対して、これらを教師あり学習問題の入力として扱うために、以下の前処理を行った。

まず、各切片の画像解像度をサイズ 1024×1024 , $3\text{nm}/\text{pixel}$ からサイズ 512×512 , $6\text{nm}/\text{pixel}$ へ変換した。この変換には average pooling を用いた。この処理によって、XY 方向の画素密度は $6\text{nm}/\text{pixel}$ となり、Z 方向の切片間隔 $30\text{nm}/\text{slice}$ の $1/5$ となった。

次に、切片間の平均輝度の違い (輝度ムラ) を軽減するために、切片ごとに画素輝度の中央値を 100 切片ぶんの輝度中央値平均に揃える処理を行った。

3.2 入力パッチの切り出し

Z 補間問題では、観測された切片画像の間に仮想的な切片を想定して、その各ピクセルの輝度を観測された上下切片における周囲パターンを用いて推定する。これを教師あり学習によって解くために、教師データとなる入出力セットを以下のように用意した。

観測された連続切片画像のうち一枚を推定対象とし、推定対象切片の各ピクセルの輝度をスカラー値出力とする。これに対応して、推定対象切片の上下 2 切片ずつ合計 4 切片のうちそれぞれサイズ 33×33 のパッチ領域を切り取ってまとめたサイズ $33 \times 33 \times 4$ の三次元パッチの輝度パターンを入力とする。このパッチ領域の切り取りは、トレーニング、テスト両方のデータセットについて 95 回ずつ行った。この処理によって得られた三次元パッチを構成する推定対象の上下 2 枚ずつの 4

切片に対応して +2 パッチ、+1 パッチ、-1 パッチ、-2 パッチと呼ぶことにする。このようにして得られた 4 枚のパッチを並べて 66×66 の画像とした (Fig 1)。

これと同様にして、X 方向、Y 方向それぞれ 5 ピクセルずつずらしながら切片全面で 95×95 回切り取りを行った。これを合計して $95 \times 95 \times 95 = 857375$ 枚集めたものを入力パッチ画像セットとする。また、各入力パッチに対応して前述した推定対象切片の中心座標 $(17, 17)$ の輝度値を教師出力信号とした。

3.3 教師あり回帰問題としての定式化

D 次元の入力ベクトル $\mathbf{x}_i$ と 1 次元の教師信号 t_i の組 $\{(\mathbf{x}_i, t_i); i = 1, \dots, N\}$ が与えられている。ここで $D = 66 \times 66$, $N = 857375$ である。このとき、教師あり回帰問題として Z 補間問題を解くことを考える。

教師あり回帰問題では、パラメトリックな関数 $y(\mathbf{x}, \mathbf{w})$ を考え、回帰誤差

$$E(\mathbf{w}) = \sum_{i=1}^N \|y(\mathbf{x}_i, \mathbf{w}) - t_i\|^2 \quad (1)$$

を小さくするようにパラメタ $\mathbf{w}$ を決定する。

パラメトリックな関数として、最も単純には以下の線形モデルを考えることができる。

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^D w_j x_j \quad (2)$$

ここで $\mathbf{w} = (w_1, \dots, w_D)$ は重みパラメータである。このパラメータを合計回帰誤差最小基準でもとめた場合を本稿で「線形回帰法」と呼ぶ。また非線形な方法として CNN を用いる方法を提案する。

またこれらの比較対象として、線形モデルのうえで単純な構成のもとで適当に固定する simple average (2 pixels) と simple average (18 pixels) の二つを考える。

*1 ISBI Challenges 2013,
<http://www.biomedicalimaging.org/2013/program/isbi-challenges/>

まず、simple average (2 pixels) では、推定対象の上下切片 (+1 切片と-1 切片) それぞれから推定対象ピクセル直上と直下の 1 ピクセルの輝度値を取得し、それらの平均を出力とする。

次に、simple average (18 pixels) では、+1 切片、-1 切片 それぞれから推定対象ピクセルの直上、直下のピクセルおよびその周辺 8 ピクセルの輝度値を取得し、それら合計 18 個の輝度値の平均を出力とする。

同様にして、simple average (50 pixels) も考えることができるが、表 1 に示すように、予備シミュレーションの結果では simple average (50 pixels) より simple average (18 pixels) の方が MSE の値が小さくなったので本論文では議論の枠から外す。

表 1: 実験条件 1(本文 4.1 章), 2(本文 4.2 節) の各実験条件において、左から順に simple average (2 pixels), simple average (18 pixels), simple average (50 pixels) について MSE の値を示した。また、各実験条件について最も低い MSE 値を太字で示した。

MSE	simple average (2 pixels)	simple average (18 pixels)	simple average (50 pixels)
実験条件 1	1027.8	915.5	927.6
実験条件 2	792.4	711.1	714.8

3.4 CNN モデルの構成

CNN は多層ニューラルネットワーク (DNN) のスパースな実装の一つであり、convolution 層と pooling 層の重ね合わせと重み共有によって、入力画像の幾何学的変化に対する不変性をスパースで効率的な重みパラメタ表現として獲得することができる。重みは入力画像の画素ごとではなく convolution フィルター毎に共有されているので、通常の DNN と比べて計算量が少なくなり、かつ誤差逆伝播学習による誤差勾配が発散しにくいことから過学習も起こりにくくなっている。また、DNN と同様に入力に対する非線形性を持つため、線形モデルよりはるかに高い表現能力を持つ。ここまで非線形関数としてきたものは、画像認識においては特徴抽出を表す変換であり、広く使われているものの中では SIFT や SURF といった特徴抽出子が存在する。DNN では非線形関数の学習が行えるので、隠れ層で特徴抽出を行っていると考えられることもできる。

convolution 層は、モデルに検出対象の平行移動とサイズ変化に対する不変性を持たせつつ局所特徴を抽出する機能を担う。このために適当なサイズの convolution filter を複数保持しながら、入力画像に対してそれぞれの畳込み和の結果を求め次の層に送る。この層から次の層に送られる画像群は特徴マップとも呼ばれる。

pooling 層では、convolution 層から送られてきた特徴マップを入力として、 $n \times n$ (多くの場合 $n = 2$) の領域から単一の輝度値を計算する処理を $n \times n$ の領域が重ならないように連続的に行い、非線形関数に入力することによって出力を得る。以上の処理によって出力画像の辺の大きさは $1/n$ になり、ユニット数を大幅に減少させることによってパラメータの学習を高速化し、さらにピクセル毎に独立なガウシアンノイズのようなごく小さい変化に対する出力の変化を鈍感にすることができる。プーリング時の輝度値の計算方法は、領域内の輝度値の中で最大値を出力とする max pooling や、平均値を出力する average pooling がよく用いられる。

本論文で用いる CNN の実装は全て、CNN に特化した Deep Learning のライブラリである Caffe を用いて行った。入力

$66 \times 66 = 4356$ 次元で、隠れ層一層とした。convolution 層のフィルターサイズは 11×11 、畳み込み計算時の移動単位は 1 ピクセルずつとし、特徴マップ数は 64 とした。pooling 層では max pooling をフィルターサイズ 2×2 で互いに重ならないように適用した。活性化関数として ReLU [Nair 10] を用いた。

学習に関するパラメータは、学習率を $\eta = 10^{-8}$ 、モーメント係数を $\mu = 0.9$ 、荷重減衰係数を $\lambda = 0.0005$ と設定した。batch size はトレーニング時に 64、バリデーション時に 100 とした。iteration は 1000000 としたので、トレーニング用画像 857375 枚であれば、1epoch は $857375/64 \simeq 13396$ batch となるので、 $1000000/13396 \simeq 75$ epoch 学習が行われる。上記の設定で入力画像がサイズ 66×66 の場合、Nvidia 社製 GPU (tesla m2070) を用いたシミュレーションに 20 時間程度要した。

4. 結果

本研究で用いたデータは、2 つの異なる電子顕微鏡連続切片画像ブロックからなり、神経細胞やその他血管などの組織から成る点で局所的には同様の性質を持つように見えるが、大域的には繊維組織の走方向の偏りなどがブロック間で異なる。また、偏りはブロック内でも近い切片同士で類似しやすい傾向がみられる。そこで、実験条件 1 ではパラメータのトレーニングとテストとで異なるブロックから教師ありデータセットを用意し、実験条件 2, 3 ではそれぞれ共通ブロック内、近傍切片内で共通にトレーニング・テストデータを用意して性能比較を行った。

4.1 実験条件 1

実験条件 1 のもとで simple average (2 pixels), simple average (18 pixels), 線形回帰, CNN, それぞれによって Z 補間を行った結果を、図 2 に示した。補間誤差の定量評価として、出力画像と真の画像の間の平均二乗誤差 (MSE) と spearman の順位相関係数をそれぞれ表 2 と表 3 に示した。

表 2 を見ると、simple average (18 pixels) で MSE が最も低くなり、CNN では simple average (18 pixels) に比べて MSE がやや高くなった。表 3 でも同様に、simple average (18 pixels) が最も高くなっている。以上より、simple average (18 pixels) が最も良い補間となっていることがわかる。

また、図 2 を見ると、推定に用いたピクセルの数が線形回帰, CNN とくらべて少ない simple average (2 pixels), simple average (18 pixels) の方がくっきりした画像が得られているように見える。

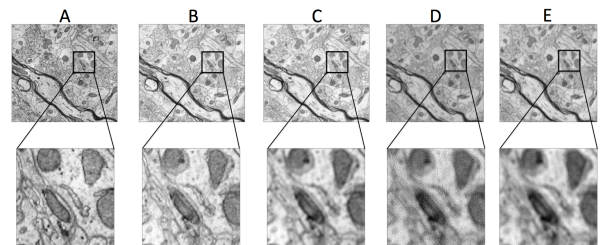


図 2: 実験条件 1(本文 4.1 節) における補間結果。A は真の画像, B は simple average (2 pixel), C は simple average (18 pixels), D は線形回帰, E は CNN に対応する。それぞれ上段に補間結果、下段に一部 100×100 pixel ぶんを拡大した画像を示した。

表 2: 実験条件 1(本文 4.1 節), 2(本文 4.2 節), 3(本文 4.2 節) の各実験条件において、左から順に simple average (2 pixels), simple average (18 pixels), 線形回帰, CNN について MSE の値を示した。また、各実験条件について最も低い MSE 値を太字で示した。

MSE	simple average (2 pixels)	simple average (18 pixels)	線形回帰	CNN
実験条件 1	1027.8	915.5	981.8	928.9
実験条件 2	1027.8	915.5	909.7	908.2
実験条件 3	792.4	711.1	712.7	705.6

表 3: 実験条件 1(本文 4.1 節), 2(本文 4.2 節), 3(本文 4.2 節) の各実験条件において、左から順に simple average (2 pixels), simple average (18 pixels), 線形回帰, CNN について spearman の順位相関係数の値を示した。また、各実験条件において、最も高い spearman の順位相関係数の値を太字で示した。

spearman rank correlation	simple average (2 pixels)	simple average (18 pixels)	線形回帰	CNN
実験条件 1	0.606	0.646	0.658	0.663
実験条件 2	0.606	0.646	0.669	0.679
実験条件 3	0.725	0.751	0.787	0.789

4.2 実験条件 2,3

トレーニング用とテスト用データセットの出処となるブロック間に構造特徴の偏りが想定されている。この偏りが性能に与える影響を見るために、以下の実験条件 2,3 で同様の比較を行った。

実験条件 2 では、実験条件 1 においてテストに用いた 100 切片からなるブロックのうち、80 切片をトレーニング用、20 枚をテスト用とした。学習結果の評価には、実験条件 1 と同じ画像を用いた。実験条件 3 では、学習結果の評価に上下切片との輝度の差が比較的小さく推定が容易と思われる画像を特に選んで用いた。

実験条件 1,2,3 の比較を表 2 と表 3 に示した。表 2 によれば、実験条件 2 での MSE の値は線形回帰, CNN 両方で実験条件 1 との比較で向上しており、一方固定された重みの値を用いる simple average (2 pixels) と simple average (18 pixels) では MSE の値に変化はない。結果として線形回帰と CNN がほぼ同じ値で最低値となった。また、表 3 を見ると、CNN が最も順位相関係数が高くなり、線形回帰と simple average (18 pixels) では近い値が出ているがやや simple average (18 pixels) における順位相関係数の値の方が高くなっている。実験条件 3 でも同様の傾向が見られた。以上の結果により、連続切片を推定する問題においては CNN が最も精度の高い補間を実現した。

5. 結論

電子顕微鏡連続切片画像における Z 方向補間問題の定式化を行った。線形モデルおよび CNN を用いた手法を提案し、各手法の比較検討を行った結果、MSE と spearman の順位相関係数による評価において CNN が高い性能を持つことを示した。CNN による補間結果から、神経線維の走行構造パターンに関する情報抽出学習に至っていないことが示唆された。今後は、学習データを増やすことと CNN の層構造の工夫による補間性能の改善を目指す。神経線維の三次元走行構造再構成の自

動化に対する貢献は今後の課題である。

参考文献

- [Hayworth 06] K. J. Hayworth, N. Kasthuri, R. Schalek, and J. W. Lichtman.: Automating the collection of ultrathin serial sections for large volume term reconstructions, *Microscopy and Microanalysis*, vol.12, no. S02, 86-87, 2006.
- [Kaynig 13] V. Kaynig, A. Vazquez-Reina, S. Knowles-Barley, M. Roberts, T. R. Jones, N. Kasthuri, E. Miller, J. Lichtman, and H. Pfister.: Large-Scale Automatic Reconstruction of Neuronal Process from Electron Microscopy Images, arXiv:1303.7186, 2013.
- [LeCun 98] Y. LeCun, L. Bottou, Y. Benjio, and P. Haffner.: Gradient-based learning applied to document recognition, *proceedings of the IEEE*, 86(11), 2278-2324, 1998.
- [Lichtman 08] J. W. Lichtman and J. R. Sanes.: Ome sweet ome: what can the genome tell us about the connectome?, *Current opinion in neurobiology*, vol. 18, no. 3, 346-53, 2008.
- [Helmstaedter 11] M. Helmstaedter, K. L. Briggman, and W. Denk.: High-accuracy neurite reconstruction for high-throughput neuroanatomy, *Nature neuroscience*, vol. 14, no. 8, 1081-8, 2011.
- [Chklovskii 10] D. B. Chklovskii, S. Vitaladevuni, and L. K. Scheffer.: Semi-automated reconstruction of neural circuits using electron microscopy, *Current opinion in neurobiology*, vol. 20, no. 5, 667-75, 2010.
- [Turaga 12] S. C. Turaga, J. F. Murray, V. Jain, F. Roth, M. Helmstaedter, K. Briggman, W. Denk, and H. S. Seung.: Convolutional networks can learn to generate affinity graphs for image segmentation, *Neural computation*, vol. 22, no. 2, 511-38, 2012.
- [Funke 12] J. Funke, B. Andres, and F. A. Hamprecht.: Efficient automatic 3D-reconstruction of branching neurons from EM data, *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1004-1011, 2012.
- [LeCun 89] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, et al.: Backpropagation applied to handwritten zip code recognition, *Neural Computation*, 1, 541-551, 1989.
- [Nair 10] V. Nair, G. E. Hinton.: Rectified linear units improve restricted Boltzmann machines, *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 807-814, 2010.